

# On the symmetry approach to quantum mechanics and some relations to ordinary statistics

Inge S. Helland  
Department of Mathematics  
University of Oslo  
P.O.Box 1053 Blindern  
N-0316 Oslo  
Norway

Running title: Symmetry approach to quantum mechanics.

## Abstract

Taking a recent article by A. Bohr and Ulfbeck (1995) as a point of departure, it is described how quantum mechanics can be founded on the representation theory of symmetry groups. Some discussion of this theory is made, and an example based upon the permutation group for three objects is described in detail. The conclusion seems to be that the quantum theory as such can not be limited to the atomic and subatomic world. The theory is of relevance whenever one has a probability model that is derived from a nonabelian symmetry group and when there in addition is some linear structure between parameters or between observators of the system. Some areas of statistics where this relationship can be of importance, are briefly described.

Keywords and phrases: Foundation of quantum theory; group theory; linear models; non-commutative groups; probability; quantum mechanics; representation theory; statistical modelling; symmetry.

## 1 Introduction

Since the beginning of this century we have lived with a situation where the concept of probability apparently has meant one thing for statisticians, probabilists, actuarians, model makers in biology and economics etc. and a slightly different thing for nuclear

physicists, elementary particle physicists, quantum chemists etc.. The latter have used the by now well known, but fairly abstract quantum theory, while the former have based their reasoning on Kolmogorov's axioms, saying that probabilities are normed measures on a measurable space. A synthesis of the two worlds has relatively recently been made in non-commutative probability theory (Meyer, 1993), but this is even more abstract and inaccessible than ordinary quantum mechanics. Also, there seems to be a general feeling that the increased level of abstraction has not gone hand in hand with new results that can be tested by experiments. Part of the reason for this may be that abstraction is difficult to combine with intuition, at least for most people, and intuition is a necessary ingredient if science shall proceed in new directions.

It should be emphasized that the outcome of any single experiment with a fixed experimental arrangement can always be described by using the ordinary probability concept, also in the world of particle physics. It is in cases where several arrangements are possible, as when one has the choice between measuring position or momentum of a particle, that quantum mechanics gives results that can not be reached by ordinary probability theory. In addition, quantum theory gives definite rules for *computing* probabilities, also in cases where the ordinary probability concept can be used in principle.

An interesting new approach to quantum mechanics based upon the symmetry concept and on group theory has recently been proposed by Aage Bohr and Ole Ulfbeck (Bohr and Ulfbeck, 1995). Their main thesis is that what they call primary manifestation of symmetry is the basic cause of quantal indeterminacy. While high energy physicists have used group theory as one of the most important supply for their tool kit for decades now, the aim of Bohr and Ulfbeck has been to propose symmetry groups as the *only* tools at the quantum level. This is a radical thought, even though it does not seem to be revolutionary to physicists familiar to group theory. The theory is only incompletely developed in the 1995 paper, but the idea is definitively worth investigating further.

The main thesis that probability models on the basic level of elementary particles should be based upon symmetry, is not difficult to accept, and the road from symmetry to group theory is also quite well known. When the symmetry group is commutative (abelian), it seems to be fairly obvious how the corresponding probability measure should be: Equal probability of each element in the finite case and Haar measure in general. In the non-abelian case it is known that every finite group has irreducible unitary representations of dimension more than 1. In the theory of Bohr and Ulfbeck it is the eigenvectors of these matrices that constitute the state vectors of the system; the matrices themselves are of the form  $\exp(iA)$ , where  $A$  are observators of the system.

One purpose of this paper is to give illustrations of this mechanism in simple

settings, and we will also give a sketch of the theory itself. We start by a brief recapitulation of ordinary quantum theory. After that we give the necessary material from group theory and representation theory. It is then described, partly with reference to the well-known Aspect experiment, how the representation theory of groups can be used to give a foundation of quantum theory. We limit ourselves to finite groups, or more generally to compact groups in this paper. An example based upon the permutation group of 3 objects is described in detail. In the last part we will discuss briefly some possible links to the statistical theory outside the quantum world.

I have attempted to include - albeit somewhat sketchy - the necessary background material from physics and mathematics. The paper is intended for statisticians with interest in the field, but it can also be read by interested mathematicians or mathematical physicists.

## 2 Elements of traditional quantum mechanics.

Quantum mechanics was developed in the first part of this century by pioneers like Planck, Heisenberg, Schrödinger and Bohr. This development was extended and summarized by Dirac (1930) and given a firm mathematical footing in von Neumann (1955), which is the basis of widely used textbooks like Messiah (1969). A more modern approach to the connection with ordinary probability theory is non-commutative quantum probability as discussed in Meyer (1993). I will not discuss this approach in any detail in this paper, since it is firmly based on the axioms of von Neumann, and we will take another route here. Nevertheless it is useful to recapitulate here the main elements of ordinary quantum mechanics.

The states of a physical system are defined as normed vectors in a complex Hilbert space. (For the simple systems that we will discuss below, it is enough to take a finite-dimensional space as a point of departure.) The probability amplitude for transition between two states is the inner product between the corresponding vectors, and the probability of transition is the squared norm of this amplitude.

Mixed states are in general defined as convex combinations of one-dimensional projections upon the state vectors.

The observables of the system are identified with selfadjoint operators  $A$  on the Hilbert space, i.e. those that have real eigenvalues; we will limit ourselves to operators with a finite set of eigenvalues here. These eigenvalues are the possible observed values of  $A$ ; if the state is an eigenstate of  $A$ , then the corresponding eigenvalue is observed with probability 1; in general the expectation of this observed value is the trace of  $A\rho$ , where  $\rho$  is the state projection. If the state of the system is

not a pure eigenstate, then the outcome of a measurement will always be random; the probability distribution can be found by decomposing into eigenstates.

The time development of states in (non-relativistic) quantum systems are described by the Schrödinger equation involving the Hamilton operator of the system. However, there is another curious way that states can change, namely by a measurement process. Once a measurement of an observable  $A$  is done, the state is changed into an eigenstate of  $A$ . Although at first sight unfamiliar to statisticians, this is not far from a well known situation in ordinary statistics where the probability distribution of a variable changes after an observation is made, usually into a degenerate distribution.

However, in quantum mechanics the corresponding situation has been considered by many to be in conflict with an assumption that the state vectors must be assumed to have some independent existence. In some sense the situation was saved by a many-world theory taking the relative state formulation of Everett (1957) as a point of departure, but a theory involving many parallel worlds existing hidden from each other at any time, seems very difficult to accept.

An alternative seems to be that these state vectors must be considered merely as tools, and that there is another theory behind everything that can be used to justify the calculations described above. Several attempts of such theories are listed in Gudder (1988).

### 3 A sketch of the Aspect experiment

As described for instance by Kümmerer and Maassen (1996), a Calcium atom excited by a laser can emit pairs of photons in opposite directions with opposite polarization. Now the (linear) polarization for any such photon can be measured in any direction orthogonal to the beam by means of a polarization filter. The result of such a measurement can be looked upon as a random variable which takes the value 1 if the photon passes through the filter, 0 otherwise. If the filters have the same direction for both photons in the pair emitted from the Calcium atom, the only possible results (ideally) are of the form 0 for one filter and 1 for the other filter.

Similar experiments have also been proposed and performed with spin of particles. Experiments essentially of this kind (but with position and momentum instead of polarization) were proposed by Einstein et al. (1935) in an attempt to demonstrate the incompleteness of quantum mechanics. Several experiments of this type have now been made (see Aspect et al., 1982 and references there), and there seems to be consensus among physicists that the predictions of quantum mechanics have been confirmed in every essential way.

If polarization measurements are done for some relative measurement directions once for each photon, this will give an ordinary bivariate distribution that can be calculated from the rules of quantum mechanics. Also, a joint distribution can be found for two successive measurements done for each photon. However, the difficulty starts if we want to give concrete interpretations to these pairs of measurements in the sense that a photon should at each time *have a definite polarization direction* or at least one should be able to identify some physical variable connected to the photon which causes the different polarization measurements. If we insist on being free to adjust the relative measurement directions and do not want solutions with seemingly action at a distance, this turns out to be impossible, a problem often referred to by saying that there are no *hidden variables* behind the measurements.

This may be illuminated by using Bell's inequality (Bell, 1966, also discussed in several later papers and books). (Gill and Barndorff-Nielsen (1996) give a thorough discussion of this and other aspects from the point of view of statisticians.) For some combinations of directions the predictions of quantum mechanics lead to values of the probabilities that are inconsistent with assuming of a joint 2-variable distribution with the properties above.

More explicitly: Assume three experiments and three 01-variables  $Z_1, Z_2$  and  $Z_3$ , one variable for each experiment. If these three can be defined on a common probability space, Bell's 3-variable inequality says that one always has

$$P[Z_1 = 1, Z_3 = 0] \leq P[Z_1 = 1, Z_2 = 0] + P[Z_2 = 1, Z_3 = 0] \quad (1)$$

Letting polarization experiments be done at different relative angles and letting the  $Z$ -variables take the value 1 when a photon passes through a given filter, quantum mechanics predicts violation of these inequality for certain choices of angles.

The measurements by Aspect et al. (1982) confirm the predictions of quantum mechanics in this setting. This was hardly a surprise to physicists, who have seen it confirmed in a large number of areas and in a range of situations. From the point of view of statisticians it is not uncommon that we have to use different probability spaces when we perform different experiments. If the experimentalist is free to perform adjustments between the experiments, it may be impossible to join the different probability spaces into one.

## 4 From the theory of finite groups and their representations

Group elements can be regarded as abstract elements, or may be interpreted as transformations of some system, a simple example being permutations of a finite set.

Group elements that carry a geometrical object into another object similar to the first one describe symmetries of the object.

A finite group  $G$  has elements  $g, h, \dots$  obeying the rules:

1. There is an associative multiplication  $gh$  defined on the group elements, i.e.,  $(gh)k = g(hk)$  for all  $g, h, k$ .
2. The group contains a unit element  $e$  satisfying  $ge = eg = g$  for all  $g$ .
3. Every element  $g$  has an inverse  $g^{-1}$  such that  $gg^{-1} = g^{-1}g = e$ .

A group is called finite if the number of elements  $|G|$ , called the *order* of the group is finite. If  $gh = hg$  for all  $g$  and  $h$ , the group is called commutative or abelian, otherwise non-commutative. Two elements  $g_1$  and  $g_2$  are said to belong to the same *conjugate class* or just *class* if  $g_2 = h^{-1}g_1h$  for some  $h \in G$ . This is easily seen to be an equivalence relation; thus any given group  $G$  will be divided into a number  $k_G$  of disjoint classes. The unit element  $e$  constitutes a class by itself.

Cayley's Theorem, which is proved by looking at the general structure of a group multiplication table, says that a group of order  $|G|$  always can be regarded as (is isomorphic to) a subgroup of the group of permutations of  $|G|$  elements. Thus all finite groups can in this sense be looked upon as permutation groups.

Nonsingular real or complex matrices will also be group elements under matrix multiplication; and such representation of groups will be of particular interest in this paper. We start by observing that any group of finite order  $|G|$  is isomorphic to a certain group of  $|G| \times |G|$  matrices. This is found by regarding the set of formal expressions

$$\sum \alpha_i g_i,$$

where  $g_i$  range over  $G$  and where the  $\alpha$ 's are scalars, as a vector space of dimension  $|G|$ . The multiplication from the group is carried over in an obvious way to these formal expressions. Since this will give a linear operation, multiplication to the left by a group element  $g$  can equally well be described as a multiplication by a matrix  $M(g)$ . In this sense  $g$  is represented by  $M(g)$ .

In general, a matrix *representation* of a group is defined as a function  $D$  from the group to a set of matrices satisfying

$$D(gh) = D(g)D(h)$$

for all  $g$  and  $h$ . Then automatically  $D(e) = I$  and  $D(g^{-1}) = D(g)^{-1}$ , and it is easy to see that the matrices involved will constitute a group under multiplication. The full representation described above of a group by  $|G| \times |G|$  matrices is called the *regular representation* of the group. Any representation  $D(g)$  and any fixed nonsingular

matrix  $K$  of the same size can be used to construct another representation  $U(g) = KD(g)K^{-1}$ . If this  $U(g)$  should happen to be of the same block diagonal form for each  $g$ , then each block will constitute a new representation.

One always has the trivial representation  $U(g) \equiv I$ . The group of permutations of  $n$  objects has a representation  $P(\cdot)$  of degree  $n$  consisting of permutation matrices; each row and each column has one entry equal to 1 and  $n-1$  entries equal to 0, so that  $P(g)v$  has the same coordinates as  $v$ , but permuted according to the permutation  $g$ . This *permutation representation* is carried over to subgroups.

The similarity transformations of the type  $KD(g)K^{-1}$  are particularly useful if we take the regular representation as a point of departure; the transformed representations will also be isomorphic to the original group, and can therefore be used to describe the whole group in another way. To be able to simplify the matrix group as much as possible, it is useful to employ the complex number field, which we will do from now on in the matrix representations. Then by a similarity transformation the regular representation can be transformed to a representation  $U(g)$  of block diagonal form with minimal blocks  $U_r(g)$ . The following can be proved for this new representation:

1. Each  $U(g)$  can be assumed unitary (meaning  $U^\dagger U = I$ , where  $\dagger$  denotes complex conjugation plus transposition).
2. If the group is abelian, each  $U(g)$  can be taken to be diagonal.
3. In general  $U(g)$  can be taken block diagonal in a minimal way, with unitary blocks  $U_r(g)$  of size  $d_r \times d_r$ , and these can be taken as repeated  $d_r$  times.
4. The dimensions and the group order are connected by  $\sum d_r^2 = |G|$ .
5. The dimensions  $d_r$  divide the group order  $|G|$ .
6. The number of nonequivalent blocks equals the number  $k_G$  of classes in the group (i.e.,  $r = 1, \dots, k_G$ ).

The repetition of blocks that we get here for the noncommutative case, does of course not give us any extra information. The representation obtained when omitting these repetitions, is called the canonical representation (Serre, 1977). The representation given by the single minimal blocks are called the *irreducible representations*.

Note in connection to the first two points here that for abelian groups the unitary  $1 \times 1$  'matrices' on the diagonal must be of the form  $\exp(i\alpha_g)$  for each given  $g$ , implying that the same term corresponding to  $g^n$  then will be  $\exp(i\alpha_g n)$ . In the abelian case we can therefore use this to get something similar to a Fourier transform from the group elements. The general case is more complicated, but in the same

spirit. For any function  $f$  on the group elements, we can define a Fourier transform  $\hat{f}$ , whose arguments are the blocks of the irreducible representations:

$$\hat{f}(U_r) = \sum_{g \in G} f(g) U_r(g). \quad (2)$$

This transform can be inverted (Diaconis, 1988)

$$f(g) = \frac{1}{|G|} \sum_{r=1}^{k_G} d_r \text{Tr}(U_r(g^{-1}) \hat{f}(U_r)), \quad (3)$$

and there is also a Plancherel formula

$$\sum_g f(g^{-1}) h(g) = \frac{1}{|G|} \sum_r d_r \text{Tr}(\hat{f}(U_r) \hat{h}(U_r)). \quad (4)$$

The final concept that should be mentioned, is the *character* of a group representation, defined simply as the trace of the representation

$$\chi(g) = \text{Tr}(U(g)),$$

which are complex-valued functions of the group elements. Of special importance are the *irreducible characters*  $\chi_r(g) = \text{Tr}(U_r(g))$ . These are simpler to find than the irreducible representations themselves and can be used as tools to find these. It is easy to see that representations that are equivalent under the similarity transformation described above give the same characters and that two group elements  $g$  and  $hgh^{-1}$  belonging to the same class have the same values for any characters. The irreducible characters satisfy the orthogonality relations

$$\frac{1}{|G|} \sum_g \chi_r(g)^* \chi_{r'}(g) = \delta_{rr'}. \quad (5)$$

By defining the scalar product

$$\langle \psi, \phi \rangle = \frac{1}{|G|} \sum_g \psi(g)^* \phi(g) \quad (6)$$

one can write (5) as  $\langle \chi_r, \chi_{r'} \rangle = \delta_{rr'}$ .

Note that the matrix representations of the group in principle can be constructed knowing nothing but the group multiplication table. These matrices can always be considered as linear operators on some vector space  $L$ . Subrepresentations correspond to subspaces which are invariant under the (matrix) group operations.



Let now  $L$  be the vector space of any representation  $U(g)$ . It can be shown in general that any representation can be transformed into block diagonal form with blocks corresponding to irreducible representations of the group. The orthogonal projection upon the irreducible representation space corresponding to  $U_r(g)$  can be computed as

$$P_r = \frac{d_r}{|G|} \sum_g \chi_r(g)^* U(g). \quad (7)$$

*Example: Dihedral group  $D_n$  of order  $2n$ ,  $n$  odd.*

This group gives all symmetries of a regular polygon with  $n$  sides ( $n = 3, 5$  or  $\dots$ ). In addition to the unit  $e$ , we have  $n - 1$  different rotations  $f_j = f^{j-1}$ ,  $j = 2, \dots, n$ , where  $f$  is a rotation through an angle  $2\pi/n$ . In addition the group contains all reflections through lines through one corner of the polygon and the midpoint of the opposite side. This gives  $n$  group elements  $s_j$  ( $j = 1, \dots, n$ ), which can be generated from  $s = s_1$  and  $f$  as  $s_j = s f^{j-1}$ . It is easy to see that the group elements satisfy the following relations, which also define the group:

$$f^n = 1, \quad s^2 = 1, \quad s f s = f^{-1}. \quad (8)$$

The unit element  $e = f^0$  forms a class by itself. The  $n - 1$  rotations form pairs  $f^{j-1}$  and  $f^{-j+1}$ , which can be transformed into each other by any reflection. Finally, all the reflections form a single class, where the transformation of class elements is done by rotations. The number of classes is therefore  $1 + (n - 1)/2 + 1 = (n + 3)/2$ .

There are two trivial, one-dimensional irreducible representation, the first with  $U_0(g) = 1$  for all  $g$ , and the second with  $U_A(f_j) = 1$  and  $U_A(s_j) = -1$ . As one-dimensional representations these are equal to their characters. The remaining  $(n - 1)/2$  representations are all two-dimensional. They have the form

$$\begin{aligned} U_k(f_j) &= \begin{pmatrix} e^{2\pi i \cdot (j-1)k/n} & 0 \\ 0 & e^{-2\pi i \cdot (j-1)k/n} \end{pmatrix} \\ U_k(s_j) &= \begin{pmatrix} 0 & e^{-2\pi i \cdot (j-1)k/n} \\ e^{2\pi i \cdot (j-1)k/n} & 0 \end{pmatrix} \end{aligned} \quad (9)$$

for  $k = 1, 2, \dots, (n - 1)/2$ . The corresponding characters are

$$\chi_k(f_j) = 2 \cos(2\pi(j - 1)k/n), \quad \chi_k(s_j) = 0.$$

General, but fairly abstract references to the theory of group representations are Serre (1977) and James and Liebeck (1993); a more elementary, but perhaps a little

oldfashioned book is Hamermesh (1962). Applications in statistics are discussed in Diaconis (1988). Applications in high energy physics are discussed in several books, for instance Müller (1989). An early account of the use of group theory in quantum mechanics is Wigner (1931).

## 5 Symmetry model for the Aspect experiment

Much of the representation theory of finite groups described in the last section can be carried over almost immediately to a large class of continuous groups, namely the so-called compact groups (Serre, 1977). These are groups endowed with a topology such that the product  $g \cdot h$  and the inverse  $g^{-1}$  are continuous, and such that the whole group is a compact space. Each such group has a *Haar measure*  $\mu$ , which is a probability measure and satisfies the invariance property  $\mu(gA) = \mu(Ag) = \mu(A)$  for measurable sets  $A$ . The concept of matrix representation carries over, and the representations are in many cases as simple as in the finite case. Several of the formulas of the previous section can be translated by replacing sums over  $g$  by integrals with respect to the Haar measure. We will limit ourselves to a simple and central example, an immediate generalization of the dihedral group that we just treated.

*Example: Rotation and reflection group  $D_\infty$ .*

This is the group of rotations and reflections of the plane which fix the origin. It can be looked upon as the limit of the dihedral group of order  $2n$  when  $n$  tends to infinity. Hence it contains the rotation  $f_\alpha = (f_1)^\alpha$  by an angle  $\alpha$  around the origin ( $0 \leq \alpha < 2\pi$ ) and the reflections  $s_\alpha = s_0 f_\alpha$  for  $0 \leq \alpha < 2\pi$ . Note that  $s_\alpha$  gives a reflection through a line tilted an angle  $\alpha/2$  with respect to the line of the reflection  $s_0$ , a fact that is easily shown geometrically. The Haar measure has the simple form  $d\alpha/4\pi$ . The two trivial representations are the same as in the finite case, and the irreducible twodimensional representations are given by

$$U_k(f_\alpha) = \begin{pmatrix} e^{ik\alpha} & 0 \\ 0 & e^{-ik\alpha} \end{pmatrix}, \quad U_k(s_\alpha) = \begin{pmatrix} 0 & e^{-ik\alpha} \\ e^{ik\alpha} & 0 \end{pmatrix}, \quad (10)$$

where  $k$  now is any positive integer. These are all the irreducible representations of the group  $D_\infty$ , so jointly the set of representations is isomorphic to the group itself. (This will be strengthened below.) The characters are

$$\chi_k(f_\alpha) = 2 \cos k\alpha, \quad \chi_k(s_\alpha) = 0$$

Bohr and Ulfbeck (1995) use translation and reflection in space as their main example. We will use a simplified version of the Aspect experiment as our example, hoping to illustrate in this way the main ideas of the symmetry approach to quantum mechanics.

Consider then a single photon moving along a straight line trajectory towards a polarization measurement apparatus which measures the polarization of the photon in one particular direction perpendicular to the trajectory. Concentrating upon the situation before this apparatus is reached, there are obvious symmetries of the photon in a plane perpendicular to the trajectory. Using ordinary geometry it is not difficult to convince oneself that the natural symmetries are just those of the rotation and reflection group described above. One particular consequence of this symmetry is that the polarization values 0 and 1, as measured by the apparatus, have equal probability, since an apparatus orthogonal to the original one must give a similar result. Furthermore, we assume that the apparatus selects one particular state (say, polarization = 1; the concept of state will be further discussed below). Then the probability that the polarization will be 1 at a second apparatus, tilted at an angle  $\alpha_0$  relative to the first one, is the quantity we will be interested in.

Whatever 'state' means, the possible states of the system are assumed to be determined by the symmetries, hence by the group structure and possibly also by the measuring apparatus. Since all relevant information about the group is contained in the matrix representation (10), it must be possible to define the state from this representation. In the present case it turns out that it is enough to look at the representation for  $k = 1$ , that is,  $U_1(g)$ . This can be shown from the following general Theorem (James and Liebeck, 1993): An irreducible representation is faithful (isomorphic to the group itself) if the corresponding character  $\chi$  satisfies  $\chi(g) \neq \chi(e)$  whenever  $g \neq e$ . This property is easy to check for the group  $D_\infty$  and the representation  $U_1$ . A direct proof can also be given.

Let the group elements be defined such that  $s = s_0$  corresponds to reflection in the direction of the first apparatus. Then it is obvious from the situation that the matrix  $U_1(s)$  is important. The eigenvalues of this matrix,  $\pm 1$ , can tentatively be put into correspondance with the two possible measurement values. A general argument that the eigenvalues of the representation must be the important quantity is the following: Unitary transformations of the form  $K^\dagger U(g) K$ ,  $K^\dagger K = I$  always give equivalent representations. The matrix elements will generally change quite dramatically under such transformations. However, the eigenvalues of the matrices remain the same.

Tentatively one can then argue also that the corresponding eigenvector (here  $\psi^{(1)} = 2^{-1/2} \begin{pmatrix} 1 & 1 \end{pmatrix}'$  if the eigenvalue is +1) gives a description of the state of

the particle after the first apparatus is passed. In that case one can also argue that the state after the second apparatus has shown a value +1 is given by the eigenvector  $\psi^{(2)} = 2^{-1/2} (e^{-i\alpha_0} \ e^{i\alpha_0})'$ . This is the eigenvector of  $U_1(s_{2\alpha_0})$ , which is the relevant matrix, since  $s_{2\alpha_0}$  is reflection through a line tilted an angle  $\alpha_0$  with respect to the line of the first apparatus.

At this point it becomes difficult to pretend that one does not know conventional quantum mechanics, but it can also be argued on general grounds that the scalar product  $\psi^{(1)\dagger}\psi^{(2)}$  is important since this quantity too is invariant under unitary transformations. Having established this, the final point is to convince oneself that the probability that the second apparatus shows spin component +1 when the the first one shows this value is

$$|\psi^{(1)\dagger}\psi^{(2)}|^2 = \cos^2(\alpha_0). \quad (11)$$

This is the usual way to calculate probabilities in quantum mechanics, and the formula itself is the wellknown one which caused difficulties relative to Bell's inequality (1) if the experiment is repeated with angles  $\beta_0$  and  $\alpha_0 + \beta_0$ .

## 6 On deriving quantum theory from representation theory

Bohr and Ulfbeck (1995) give the following heuristic argument leading among other things to the formula (11) for calculating probabilities. (We give the argument for finite groups; the generalization to compact groups is immediate). From (7) restricted to a single irreducible representation  $U_1$  of degree  $d_1$  we get

$$\frac{d_1}{|G|} \sum_h \text{Tr}(U_1(h)) U_1(h^{-1}) = I. \quad (12)$$

Thus the  $U_1(g)$ -matrices are always linearly related in this way.

Suppose one can find some meaningful way to define and interpret scalar mean values of matrices (again it is reasonable to assume that these are based upon eigenvalues). Then the mean values  $\langle U_1(g) \rangle$  can be calculated in a consistent way by  $\langle U_1(g) \rangle = \text{Tr}(U_1(g)\rho)$ , where

$$\rho = \frac{d_1}{|G|} \sum_{h \in G} \langle U_1(h) \rangle U_1(h^{-1}), \quad (13)$$

since multiplication of (13) to the left by  $U_1(g)$ , taking traces and using (12) gives the same expression on each side, which is taken as a sign of consistency. It is also

immediate that the mean value defined in this way will be invariant under unitary transformations. Having convinced ourselves that this is the right definition of mean values of representation matrices, we can take linear combinations, and find in general that, with  $\rho$  as defined above

$$\langle A \rangle = \text{Tr}(A\rho) \quad (14)$$

for all  $d \times d$  matrices  $A$ .

The equation (14) is in a sense the core of quantum mechanics. It is most commonly used for selfadjoint ( $A^\dagger = A$ ) or for unitary ( $A^\dagger A = I$ ) matrices, but as argued above, it is valid for all square matrices of the same dimension as the representation. One main function of the equation is that it can be used to identify observables with selfadjoint matrices. From means of squares of matrices, variances can be defined, and the road is short to the famous Heisenberg inequality for non-commuting observables.

In order that (14) shall give realvalued means  $\langle A \rangle$  whenever  $A$  is selfadjoint (has real eigenvalues), it is necessary that  $\rho$  is also selfadjoint, which is in fact automatic from the definition (13). One must also have  $\rho$  nonnegative definite and  $\text{Tr}(\rho) = 1$ ; both of these requirements pose explicit restrictions upon the means of the irreducible matrix elements, as will be discussed in the next section.

If the state matrix  $\rho$  satisfies  $\rho^2 = \rho$ , it necessarily has the form  $\rho = \psi\psi^\dagger$  for some state vector  $\psi$  with  $\psi^\dagger\psi = 1$ . Then  $\langle A \rangle = \psi^\dagger A \psi$ , in particular it is equal to the corresponding eigenvalue of  $A$  when  $\psi$  is an eigenvector. A further special case is when  $A$  is a projection  $\psi_\alpha\psi_\alpha^\dagger$  upon another state vector. Then we get the probabilities of transition of the type leading to equation (11).

In this paper the argument will be mostly restricted to a single irreducible representation. In principle the more complicated case with several irreducible representations present can be handled by using the general projection equation (7) together with the orthogonality relations for the group characters. In some cases, as in the simple polarization experiment example, it can be shown quite easily that it is enough to consider one representation in the sense that one representation is isomorphic to the group itself. The situation is more complicated however: One has to show that the means computed in a single representation are in one-to-one correspondence with the means computed in the regular or permutation representation. We will come back to this in section 9.

There are important situations in high energy physics (conservation of charge and of baryon number) where one under all circumstances is forced to restrict oneself to one irreducible representation or to a class of representations. These properties of matter are called superselection rules and are apparently introduced as assumptions in an ad hoc way in the standard quantum theory based upon the Hilbert space formalism. By contrast, they seem more natural in the above approach.

Most of the arguments given above are of the standard mathematical type found in all treatments of quantum theory. Mathematics itself is of course neither macroscopic nor microscopic. What is more important, is the physical foundation, here assumed to be close to the theory of symmetry groups. There is still something non-rigorous about this fundamental part as sketched above. A very important aspect, however, is that there is nothing microscopic about the arguments as such. Therefore one might hope that they have some implication also on ordinary statistical theory and practice in cases with symmetry, if the arguments are valid.

## 7 States and their potential for statistical interpretation

The arguments leading to (14) can be criticised on two counts. First, the idea to take matrices as equivalent to variables may be said to be as formalistic here as in conventional quantum mechanics. Secondly, the theory does not give any input on what the density matrix  $\rho$  should be beyond proving that (13) and (14) are consistent. All equations are satisfied for  $\rho = d^{-1}I$ , for instance.

In fact the latter solution is a very natural one for a theory based upon symmetry. When  $\rho = d^{-1}I$ , we find from (14) that  $\langle A \rangle = d^{-1}\text{Tr}(A)$ , the arithmetic mean of the eigenvalues of  $A$ , for all matrices. In general we assume that  $\rho$  corresponds to what can be called the state of the system, and  $\rho = d^{-1}I$  can in a natural way be called the *uniform state*. It is a state where all symmetries are in force. (In quantum probability it is common to use the word *chaotic state*; in a statistical context the word ‘uniform’ is more natural. It is the less exotic of all states, and the one that is most commonly used in statistical applications.)

An alternative, simple and strong argument for (14) in a group representation setting is the following (when it is valid): Assume that we have a probability distribution over the group representation basis vectors relative to some representation: This could be over the group elements in the regular representation or over the objects in the permutation representation. It seems reasonable to assume that many relevant probability distributions derivable from a group theoretic setting can be described in this way. Let  $\pi$  be a diagonal matrix with these probabilities of basis elements on the diagonal, and let  $C$  be another diagonal matrix giving the values of some random variable upon the same space of basis vectors. Then in the usual classical way the expected value of this random variable is

$$\langle C \rangle = \sum_i C_{ii} \pi_{ii} = \text{Tr}(C\pi).$$

Now transform to a minimal block-diagonal representation of the group by some unitary matrix  $K$ ; that is:  $C \rightarrow A = KCK^\dagger$ ,  $\pi \rightarrow \rho = K\pi K^\dagger$ . Then  $\text{Tr}(A\rho) = \text{Tr}(C\pi)$  is the expected value of  $A$  (or  $C$ ) in this representation. In general this will be the sum over several irreducible representations, but, as discussed above, in many cases only one of these contribute. This may be caused by conservation laws, or may be argued for by the loose statement that the irreducible components behave separately under group operations. Alternative arguments were given in the previous section, and the issue will be taken up again in section 9.

Altogether, this argument for (14) is valid for random variables that are definable in the regular or permutation representation. In general the class of ‘random variables’ that can be defined by the procedure of the previous section is larger.

In the following sections this will be examined in the simplest possible case and by comparing with related techniques and concepts that have proved valuable in various areas of statistics.

To carry out such a program in general or even in this special case, it is possible also to use the sometimes more efficient equations of quantum mechanics, but we will prefer the direct group approach here. According to equation (13) we first have to find the means of the representation matrices  $U(g)$ . The eigenvalues of at least some of these matrices must be given unequal probabilities; otherwise the uniform state will emerge. The most general solution will then be of the form

$$\langle U(g) \rangle = \sum_{j=1}^d p_j(g) \lambda_j(g) \quad (15)$$

for a  $d$ -dimensional unitary representation  $U(g)$ . Here  $\lambda_1(g), \dots, \lambda_d(g)$  are the eigenvalues of  $U(g)$ , and each  $\lambda_j(g)$  is given a probability  $p_j(g)$  such that  $\sum_j p_j(g) = 1$ . From the definition of  $\rho$  this gives

$$\rho = \frac{d}{|G|} \sum_{h \in G} \sum_{j=1}^d p_j(h) \lambda_j(h) U(h^{-1}). \quad (16)$$

## 8 Some mathematical properties of irreducible states

If  $U(\cdot) = U_r(\cdot)$  is an irreducible representation in equation (16), it gives an irreducible state  $\rho_r$ . In analogy with the corresponding uniform state, where  $\langle U_r(g) \rangle = d_r^{-1} \chi_r(g)$ , we can define in general  $\chi_r^{(p)}(g) = d_r \langle U_r(g) \rangle$ , so that

$$\rho_r = \frac{1}{|G|} \sum_{h \in G} \chi_r(h) U(h^{-1}). \quad (17)$$

The functions  $\chi_r^{(p)}$  possess some remarkable orthogonality properties relative to the ordinary irreducible characters.

**Theorem 1**

*The following orthogonality relations hold*

$$\langle \chi_r^{(p)}, \chi_s \rangle = \delta_{rs}. \quad (18)$$

Proof:

From equation (7) we get

$$\frac{d_r}{|G|} \sum_g \chi_s(g)^* U_r(g) = \delta_{rs} I.$$

Take the mean of this equation.

◇

**Corollary 1**

*The mean  $\bar{\chi}_r^{(p)}$  over conjugate classes satisfy*

$$\bar{\chi}_r^{(p)} = \chi_r.$$

Proof:

It is known from the theory of finite groups that the functions  $\chi_r$  constitute a complete orthonormal set of class functions. Since  $\bar{\chi}_r^{(p)}$  is a class function, it follows from orthogonality and completeness that  $\bar{\chi}_r^{(p)} = k\chi_r$  for some  $k$ . Since  $\langle \bar{\chi}_r^{(p)}, \chi_r \rangle = 1$ , we must have  $k = 1$ .

◇

Essentially the same proof gives:

**Corollary 2**

*If  $\text{Tr}(AU_r(\cdot))$  is a class function, then the mean  $\langle A \rangle_r$  in the irreducible state is given by the uniform mean  $d_r^{-1} \text{Tr}(A)$ .*



<i>group element</i>	<i>eigenvalue 1</i>	<i>probability</i>	<i>eigenvalue 2</i>	<i>probability</i>
$f_2$	$e^{2\pi i/3}$	$q_2$	$e^{-2\pi i/3}$	$1 - q_2$
$f_3$	$e^{-2\pi i/3}$	$q_3$	$e^{2\pi i/3}$	$1 - q_3$
$s_1$	1	$p_1$	-1	$1 - p_1$
$s_2$	1	$p_2$	-1	$1 - p_2$
$s_3$	1	$p_3$	-1	$1 - p_3$

Table 1: Representation eigenvalues and probabilities of these for the 5 non-trivial group elements in the permutation group of 3 objects.

## 9 Permutation of 3 objects: Irreducible representation and states

The simplest non-commutative group is the dihedral group with  $n = 3$ , which has order 6. This describes geometrically all symmetries of an equilateral triangle, and is therefore also equivalent to the group  $S_3$  of permutations of 3 objects. The group elements are the identity  $e$ , two rotations  $f_2, f_3$  and three reflections  $s_1, s_2, s_3$ . The multiplication table of the group is easily computed from the relations (8). The group has three conjugacy classes:  $\{e\}$ ,  $\{f_2, f_3\}$  and  $\{s_1, s_2, s_3\}$ . Correspondingly, there are three irreducible characters: (1)  $\chi_0(g) \equiv 1$ ; (2)  $\chi_A(g) = 1$  for  $g = e, f_2$  and  $f_3$  and  $\chi_A(g) = -1$  otherwise; finally (3) the most interesting character  $\chi_1$  defined by  $\chi_1(e) = 2$ ,  $\chi_1(f_2) = \chi_1(f_3) = -1$  and  $\chi_1(s_1) = \chi_1(s_2) = \chi_1(s_3) = 0$ .

This last character also gives the most interesting irreducible representation, the only one of dimension greater than 1, namely

$$U_1(f_j) = \begin{pmatrix} e^{2\pi i \cdot (j-1)/3} & 0 \\ 0 & e^{-2\pi i \cdot (j-1)/3} \end{pmatrix}, \quad U_1(s_j) = \begin{pmatrix} 0 & e^{-2\pi i \cdot (j-1)/3} \\ e^{2\pi i \cdot (j-1)/3} & 0 \end{pmatrix}, \quad (19)$$

found from equation (9). Here  $j$  runs from 1 to 3, and we have for notational convenience identified  $e$  with an element  $f_1$ . Also, note that  $e^{4\pi i/3}$  equals  $e^{-2\pi i/3}$ .

As described in the previous section, the states are determined by the eigenvalues of the representation matrices; we only develop this here for the irreducible representation with index 1, i.e., given by (19). For this we find the eigenvalues and eigenvectors, and also define the probabilities corresponding to these; eigenvalues and probabilities given in Table 1.

The mean values  $\langle U_1(g) \rangle$  are found directly from Table 1. A straightforward calculation using equations (16) and (19) then gives the state matrix  $\rho$ . A first requirement is that the trace of this matrix shall be one. This results in:

**Requirement 1:**  $q_2 = q_3 (\equiv q)$ .

The calculations then give

**Theorem 2**

*The state matrix giving all possible states for a system of 3 objects occupying 3 positions is*

$$\rho = \begin{pmatrix} q & a + ib \\ a - ib & 1 - q \end{pmatrix}, \quad (20)$$

where  $a = \frac{1}{3}(2p_1 - p_2 - p_3)$  and  $b = \frac{\sqrt{3}}{3}(p_2 - p_3)$ .

It is interesting that only contrasts in the  $p_j$ 's occur in the formula for the state matrix  $\rho$ . Therefore the next result is not quite necessary for the following, but it does follow from Corollary 1 of the previous section that the mean of  $2p_j - 1$  is zero, hence

**Requirement 2:**  $\bar{p} = \frac{1}{2}$ .

It is also of some interest that the two contrasts given in  $a$  and  $b$  are orthogonal. Using this, we find the next requirement from demanding that the state matrix should be positive definite ( $a^2 + b^2 \leq q(1 - q)$ ):

**Requirement 3:**  $\frac{2}{3} \sum_{j=1}^3 (p_j - \bar{p})^2 \leq q(1 - q)$ .

The bordering case here corresponds to the pure states, so

**Corollary 3**

*The state is a pure state if and only if  $\frac{2}{3} \sum_{j=1}^3 (p_j - \bar{p})^2 = q(1 - q)$ . The corresponding state vector is*

$$\psi = \begin{pmatrix} \sqrt{q} \\ \sqrt{1 - q} e^{-i\theta} \end{pmatrix}, \quad (21)$$

where  $\theta = \arctan b/a$ .

The other extreme is the uniform state corresponding to  $\rho = \frac{1}{2}I$  and  $p_1 = p_2 = p_3 = q = \frac{1}{2}$ .

At this point it may be of interest to insert a few remarks from ordinary quantum mechanics: The state matrix (20) is the same statevector that is used to describe spin  $\frac{1}{2}$  system and other two-dimensional systems. It is often decomposed into the so-called Pauli spin matrices, and in connection to this it is useful to write the inequality of requirement 2 as  $a^2 + b^2 + c^2 \leq \frac{1}{4}$ , where  $c = q - \frac{1}{2}$ , and where the bordering case corresponds to the pure states. Thus the possible states can be described as a ball, with the outer sphere giving the pure states and the center corresponds to the uniform state.

## 10 Permutation of three objects: Permutation representation and probabilities

One way of finding concrete interpretations of the parameters  $p_1$ ,  $p_2$ ,  $p_3$  and  $q$ , is to go from the blockdiagonal representation to the permutation representation and other representations. We first give a general result related to this.

### Theorem 3

*Assume that a group representation on block diagonal form*

$$U_b(g) = \begin{pmatrix} U_0(g) & 0 \\ 0 & U_1(g) \end{pmatrix} \quad (22)$$

*(such that  $U_0(g)$  has dimension  $d_0 \times d_0$  and  $U_1(g)$  has dimension  $d_1 \times d_1$ ) is equivalent under a similarity transformation to another representation  $U(g)$ , i.e.,  $U_b(g) = KU(g)K^{-1}$ . Let  $\rho$  and  $\rho_b$  be the corresponding state matrices, and let  $\tau_0$  and  $\tau_1$  be the upper and lower blocks of  $\rho_b$ . Assume that  $\rho_0 = d_0^{-1}(d_0 + d_1)\tau_0 = d_0^{-1}I$  corresponds to a uniform state in this representation. Let the  $(d_0 + d_1) \times (d_0 + d_1)$  matrix  $M$  be transformed into  $M_b$  by the same transformation, and let  $M_1 = M_{b,11}$  be the lower right part of the matrix  $M_b$ . Then the mean of  $M$  can be calculated solely within representation 1 if and only if  $\text{Tr}(P_0 M) = 0$ , where  $P_0$  is the projection given by equation (7) with  $r = 0$  and  $d_0 \chi_0 = \langle U_0(g) \rangle$ . In that case*

$$\langle M \rangle = \text{Tr}(M\rho) = (d_0 + d_1)^{-1} d_1 \text{Tr}(M_1 \rho_1) \quad (23)$$

where  $\rho_1 = d_1^{-1}(d_0 + d_1)\tau_1$ .

Proof:

The relationships between  $\rho_j$  and  $\tau_j$  ( $j = 0, 1$ ) are found from the fact that traces of state matrices should always be 1 and by using equation (15) for the uniform state. Since the mean is invariant under similarity transformations, we have in an obvious notation

$$\langle M \rangle = \text{Tr}(M_b \rho_b) = \text{Tr}(M_0 \tau_0) + \text{Tr}(M_1 \tau_1)$$

because  $\rho_b$  is block diagonal. Since the state 0 is a uniform state, the first term on the righthand side vanishes if and only if  $\text{Tr}(M_0) = \text{Tr}(K(P_0 M P_0)K^{-1}) = \text{Tr}(P_0 M P_0) = 0$ .

◊

We will now consider the case when  $U$  is the permutation representation of degree 3 of the group  $S_3$ . In this situation the irreducible representation with index 0 is the trivial onedimensional representation, which together with the twodimensional representation is equivalent under a similarity transformation to the threedimensional permutation representation. The trivial representation is always in a uniform state, so this assumption is vacuous here.

Consider the concrete situation where one has 3 boxes A, B and C, each of which contains one of the objects 1, 2 or 3. As we shall see in this and in the next section, different (initial) states here can be put into correspondence with *what we know about the system*, which helps interpreting the concept of state in this situation. To arrive at such interpretations, we start by finding the probability  $\pi_1(A)$  that object 1 shall be found in box A, and this probability shall be found in terms of the parameters of the state of the twodimensional irreducible representation. The probabilities  $\pi_2(A)$  and  $\pi_3(A)$  are defined similarly. Since these probabilities add to 1, finding them constitutes a twodimensional problem, so there is some hope that the probabilities can be found in terms of the irreducible twodimensional representation.

First, the constraint among the  $\pi_j$ 's is an inhomogeneous one, hence we introduce new parameters  $\gamma_j = \pi_j(A) - \frac{1}{3}$ , so that the constraint is  $\sum \gamma_j = 0$ . Note that  $\gamma_j \equiv 0$  in the ordinary equilibrium case  $\pi_j(A) \equiv \frac{1}{3}$ . We will see below that other cases are also possible and of interest; it all depends upon what information we have about the system.

One important point from quantum mechanics has been suppressed until now: One is in principle free to choose a basis for the state matrix  $\rho$ , and it pays to choose a basis corresponding to the measurement that is to be performed. We will have in mind any experiment that includes the opening of box A. This will then give the interpretation of the basis for the matrix in equation (20).

Assume now the three-dimensional permutation representation, where the coordinate axes correspond to the three objects. Let the  $3 \times 3$  matrix  $Q$  have 1 on the upper left entry and 0 at all other places. Then  $Q$  is a diagonal matrix, and the diagonal gives the indicator corresponding to the first object. Hence  $\langle Q \rangle = \text{Tr}(Q\rho)$  is the probability  $\pi_1(A)$ . Taking the mean after subtracting  $\frac{1}{3}I$  from  $Q$  gives the parameter  $\gamma_1$ . Put  $M = Q - \frac{1}{3}I$ . The projection  $P_0$  projecting upon the subspace of constant (zero) parameter sum is  $\frac{1}{3}J$ , where  $J$  is a  $3 \times 3$  with all entries equal to 1. A simple calculation shows then that  $\text{Tr}(P_0 M) = 0$ , so we can use the formula (23) to find the mean value of  $M$ . Here  $\rho_1$  is called  $\rho$  in equation (20). The calculations then give

**Theorem 4**

*The probabilities of finding the respective objects in box A are:*

$$\pi_1(A) = \frac{2}{3}(p_1 - \bar{p}) + \frac{1}{3}, \quad \pi_2(A) = \frac{2}{3}(p_2 - \bar{p}) + \frac{1}{3}, \quad \pi_3(A) = \frac{2}{3}(p_3 - \bar{p}) + \frac{1}{3}.$$

Proof:

Let  $P(\cdot)$  be the permutation representation of the group; the permutation matrices  $P(g)$  are easy to write up. By transforming equation (13) to this three-dimensional representation, we find

$$\rho_1 = \frac{3}{6} \sum_{h \in G} \langle U_1(h) \rangle P(h^{-1}).$$

Hence

$$\gamma_1 = \frac{2}{3} \cdot \frac{1}{2} \sum_{h \in G} \langle U_1(h) \rangle \text{Tr}(Q - \frac{1}{3}I) P(h^{-1}).$$

Using equation (19) and the probabilities given in Table 1 to evaluate  $\langle U_1(h) \rangle$ , we find  $\gamma_1 = \frac{2}{3}(p_1 - \bar{p})$ .

◊

Remarks

1. Since only contrasts in the  $p_j$ 's occur in the formula for the state matrix, these parameters may be looked upon as containing an arbitrary additive constant. By Requirement 2 we do have, though,  $\bar{p} = \frac{1}{2}$ . Hence  $\pi_j(A) = \frac{2}{3}p_j$ .

2. The probabilities depend only upon the parameters  $p_j$ , not on  $q$ . Now the three diagonal matrices  $Q$  corresponding to  $j = 1, 2, 3$  span the space of matrices that are diagonal in the permutation representation, so in order to find events described by a diagonal matrix whose probabilities depend upon  $q$ , we will have to go to the regular representation, which has degree  $3! = 6$ . This can be done using the same technique, but the calculations are cumbersome. We will take the parameter  $q$  up again in the next section.

3. As demonstrated in the proof of Theorem 3, an alternative way to do the calculations leading to Theorem 2 is to apply a similarity transformation to equation (16) and do all calculations with  $3 \times 3$  matrices. The advantage with that approach is that we do not have to give explicitly the matrices of the similarity transformation.

4. The interpretation of Theorem 3 is as follows: If the system (boxes, contents, information about boxes and contents) is in a state corresponding to  $p_1, p_2, p_3$  and  $q$ , and we open box A, the probability of finding object  $j$  is  $\pi_j(A) = \frac{2}{3}p_j$ . If we open box A, the information about the system changes; hence the state changes.

## 11 Permutation of three objects: Interpretations

That such a simple system as permutation of three objects makes rise to so complicated calculations as those discussed in the previous section, may come as a big surprise to many. However, even classical probability calculations for this system may be difficult enough; it suffices to recall the confusion that arose short time ago around the 3-door-problem of Marilyn vos Savant (see Morgan et al., 1991, and references there).

The main issue from a statistical point of view is that a model builder should formulate in a precise way the information that he or she has about the system. Different information will in general lead to different models - or in the present language: different states.

Assume again that we have three boxes A, B and C, each containing one of three objects 1, 2 or 3. Any model of interest can ultimately be formulated in terms of the probabilities  $\pi_1(A)$ , the probability that the object 1 is found in box A etc.. For a moment look at this system in the usual way based on ordinary probabilities. Then one way to specify these is to give probabilities for each permutation 123, 132 etc, where the position denotes the box. If now all the marginal probabilities  $\pi_1(A)$  etc. are equal, it turns out that all probabilities of the system depend upon a single extra parameter  $\eta$ .

In Theorem 3 we saw that in the experiment where box A was opened, the probabilities  $\pi_1(A)$  etc. equal linear combinations of the quantum parameters  $p_1$  etc.

Now the state matrix of the quantum system depends linearly upon the parameters  $p_1, p_2, p_3, q$ , and therefore any computed quantum expectation/ probability will do the same. This leads to the conjecture that the parameter  $\eta$  will have a simple linear relation to the quantum parameter  $q$ . Looking at the ranges of the two parameters when the marginal probabilities are equal, natural conjectures are  $3\eta = q$  or  $3\eta = 1 - q$ . In fact this conjecture can be confirmed using more or less physical arguments.

### Theorem 5

*In addition to opening box A let our experiment consist of determining whether the permutation of objects over boxes is even or odd. Then, with a suitable basis  $q$  is the probability that the permutation is even (123, 312 or 231).*

### Proof.

Picture the possible states as a equilateral cardboard triangle lying on a table with one corner of each of the letters A, B and C, the position of each letter indicated on the table. Let one of the sides of the triangle be white, the other black. It is clear that a full specification of the state is equivalent to specifying which object is at letter A in addition to the colour of the triangle. In a state where all marginal probabilities are equal, this gives two quantities that can be measured independently of each other, and this is also true in general if we interpret ‘independent’ not in the statistical sense but as ‘can be specified by independent parameters’.

In Theorem 3 the state parameters  $p_j$  were put into one-to-one correspondence with the probabilities  $\pi_j(A)$ . The last parameter  $q$  must therefore be related to the probability of getting a white or a black triangle. Calculating probabilities from equation (20) always gives a linear expression in  $q$ , while calculating probabilities from a state specified by  $\pi_0 \equiv P(\text{black})$  and  $\pi_j(A)$ , ( $j = 1, 2, 3$ ) again gives a linear equation in  $\pi_0$ . Hence  $\pi_0$  and  $q$  must be linearly related. Since they both vary from 0 to 1, the only possibilities are  $\pi_0 = q$  and  $\pi_0 = 1 - q$ , the choice being related to the choice of basis for the irreducible representation. We will choose a basis such that

$$P(\text{black}) = P(\text{even permutation}) = \pi_0 = q.$$

◊

It is important to stress that quantum parameters do not in general have existence as such unless they are connected to an experiment where they can be measured/ estimated.

In order to understand a little more about how the (initial) state of a system shall be interpreted, we look at 3 ways that we can gain information in the 3 objects-example, so that the state will no longer be the simple uniform state. In all cases there is another person P who opens one or two boxes, and then tells the observer M everything that he knows, or just part of it. After that M calculates the (his) probability  $\pi_j \equiv \pi_j(A)$  that box A contains object  $j$  ( $j = 1, 2, 3$ ). All these probabilities can be calculated either in a classical (easy) way or by using quantum mechanics.

I) P opens box B. Then M and P enter into some information exchange process. After this, the most general statement M can make is  $P(P \text{ found object } j) = c_j$ ,  $j = 1, 2$  and  $3$ , where  $c_1$ ,  $c_2$  and  $c_3$  are fixed constants adding to 1. By a standard conditioning-argument this implies  $\pi_j(A) = \frac{1}{2}(1 - c_j)$  ( $j = 1, 2, 3$ ), and if we normalize the quantum parameters such that  $\bar{p} = \frac{1}{2}$ , it follows that  $p_j = \frac{3}{4}(1 - c_j)$ . The inequality in requirement 2 translates into

$$\sum_j (c_j - \frac{1}{3})^2 \leq \frac{3}{8}q(1 - q)$$

The lefthand side here has its maximum  $\frac{2}{3}$  in the trivial case where  $c_j = 1$  for one fixed  $j$ . The righthand side also has the same maximum, which is reached only when  $q = \frac{1}{2}$ , and this then gives that the trivial case is a pure state. In all other cases  $q$  can take values in an interval, the end points of that interval corresponding to pure states.

II) P opens both B and C. Again M and P exchange information partially, and M states for  $j = 1, 2$  and  $3$ :  $P(\text{object } j \text{ is not in } B \text{ or in } C) = d_j$ , where the  $d_j$ 's are constants adding to 1. Then  $\pi_j(A) = d_j$ ,  $j = 1, 2, 3$ , and the discussion is parallel to the one under I).

III) P opens both B and C and states for one particular  $j$  that  $x_j$  is here. Assume  $j = 2$ ; the other cases are similar. Then  $\pi_1(A) = \pi_3(A) = \frac{1}{2}$  and  $\pi_2(A) = 0$ , implying  $p_1 = p_3 = \frac{3}{4}$  and  $p_2 = 0$ . The requirement 2 inequality boils down to  $q(1 - q) \geq \frac{1}{4}$ , demanding  $q = \frac{1}{2}$  ( $= \pi_1(A)$  in agreement with the interpretation of  $q$  just given).

As these examples may illustrate, there is no disagreement between the results of quantum mechanics and classical probability when both can be used. When the latter is applicable, it usually gives the simplest calculations. The strength of quantum mechanics is that it makes us able to calculate probabilities over a larger range of cases (same state, different experiments). In all examples above, the discussion only went to the point where the state could be specified. Once the state is found, one can specify an experiment (open box A or box C, determine the permutation as even or odd etc.), and expectations and probabilities may be calculated.

The two cases I and II are interesting in that they show that a large part of the states of the form (20) can be interpreted in this simple way. Therefore a large part of



all possible spin states for spin  $\frac{1}{2}$  particles can be put into one-to-one correspondence with examples of this type. The main properties of these states as described here, is that they involve the statements of two persons; except for that they can pass as hidden variable models of the type that theoretical physicists have been debating for decades.

This two-person structure also opens for many applications in statistical inference, where much of the theory is based upon exchange of information from nature to an observer. It is not the place to go into a detailed study of this aspect here. We will only mention some areas of statistics where group-theoretical concepts of the type discussed here arise in a natural way.

## 12 Possible further links to statistics.

From a statistical point of view, about the simplest and most natural group to look at is given by random permutation of  $n$  elements. As in the case  $n = 3$  we can study this permutation group either in the permutation representation (dimension  $n$ ) or in the regular representation (dimension  $n!$ ). The latter is fairly complicated, but has a well studied structure; its partition into irreducible parts is treated in several textbooks in group theory and in elementary particle physics. The former has a very simple structure: Its decomposition into irreducible parts contains the trivial one-dimensional part and in addition an  $(n-1)$ -dimensional part, immediately generalizing the case  $n = 3$ .

For a statistician this corresponds to a single random sample. The one-dimensional irreducible representation reflects the mean structure, the other irreducible part reflects the variation around the mean. It is of course well known that the estimation of this variation leads to a sum of squares with  $n - 1$  degrees of freedom.

One way analysis of variance (with random effects) can be thought of as nesting (taking the wreath product of) two groups of this kind into each other; two way analysis of variance as crossing (taking the direct product of) two such groups with each other. Nelder (1965) introduced a class of block structures that has had great influence on the British school of experimental design, by combining these two ways of combining designs in all possible ways. More general, but related classes of design have been analyzed by Tjur (1984) and Bailey (1991), among others. Bailey (1991) concentrates on randomization theory using group theoretic methods. The irreducible components of the randomization groups are called *strata* and are of vital importance both in the theory and in the practice of experimental design. Alternatively, the strata can be looked upon as eigenspaces or invariant spaces of the (unknown) covariance matrix of the error distribution.

Many papers from this school of inference treats the error structure and the treatment structure separately, but Tjur (1984, 1991) unites the two aspects in one model. Helland (1996a) is an attempt to see this class of models from a population point of view, starting with a random model and then arriving at fixed effects by conditioning in this random model. Much more of the discussion here could have been made in a group theoretic language, and the concept of strata is of vital importance.

The usual (sample version of the) randomization theory connected to the strata is based on a uniform state. Other states are important, however, when we look at another symmetry aspect of analysis of variance models: Symmetry among the fixed effects or treatment structure. The treatment combinations that are of interest are often made explicit in a set of orthogonal *contrasts*. The freedom one has in choosing a set of orthogonal contrasts has a close connection to the freedom one has in choosing a set of orthogonal eigenvectors as (final) states in quantum mechanics. According to the philosophy of the present paper, both can be given the same group interpretation.

As for initial states, other alternatives than the uniform one seem to be relevant in cases with confounding of contrasts, like in fractional factorials. In this language, a completely pure state will not be of much interest in a designed experiment, as it will imply testing against an error with just one degree of freedom.

The other important class of normal-linear models, regression analysis, also has important aspects related to group invariance, in my opinion the most important ones are connected to prediction methods for cases with a large number of explanatory variables. In many practical cases where such methods are applied, it is natural to look upon the explanatory variables as random variables, having a covariance matrix  $\Sigma$ . In Helland (1990) it is proved that the most popular chemometric prediction method, partial least squares, in its population version can be looked upon as a method where the number of  $\Sigma$ -invariant spaces ('strata') on which the regression vector has a non-vanishing component is the quantity to keep low under model reduction. In von Rosen (1994) the invariance space aspect is further studied. In Helland and Almøy (1994) various prediction methods are compared under this model. In Helland (1996b) connections to latent variable models are discussed. It may be of some relevance that Agnar Höskuldsson for some years have written articles on the relationship between partial least squares and similar methods to quantum mechanics (; there is a book to appear, Höskuldsson, 1996).

One very controversial problem in the theory and practice of statistics is how statistical models should be conditioned. In Helland (1995a, 1995b, 1996a) I argue that it should be important for the choice of conditioning to look at the population for which the inference is intended. This was illustrated by analysis of variance examples; I could have chosen examples from contingency tables, and both these cases could have been formulated in group-theoretical terms.

Barndorff-Nielsen (1995) discusses a very interesting example of McCullagh (1992, 1993) on the choice of conditioning in a sample of size  $n$  from a scaled and centered Cauchy distribution with density

$$f(y) = \frac{1}{\pi|\theta_2|\{1 + (y - \theta_1)^2/\theta_2^2\}}.$$

Here  $-\infty < \theta_1 < \infty$  and  $\theta_2 > 0$ . The maximum likelihood estimates  $\hat{\theta}_1$  and  $\hat{\theta}_2$  can be found numerically. Arguments going back to Fisher can be used to claim that one should condition upon the ancillaries

$$a_j = \frac{y_j - \hat{\theta}_1}{\hat{\theta}_2}.$$

The problem here, however, is that the family of distributions is invariant under the transformation  $y \rightarrow 1/y$ , but with new parameters and non-equivalent ancillaries. McCulloch (1992) studies this problem by transforming the ancillaries to the unit circle of the complex plane by the transformation

$$z_j = \frac{1 + ia_j}{1 - ia_j}$$

and by showing that the ancillaries under  $y \rightarrow 1/y$  are reflected through the angle  $\arg(\theta)$ , where  $\theta = \theta_1 + i\theta_2$ .

More generally, the family of scaled and centered Cauchy distributions is invariant under the transformation  $y \rightarrow \frac{ay+b}{cy+d}$ , where  $a, b, c$  and  $d$  are real. Matrix calculus can in general be used to facilitate calculations with such bilinear transformations. Repeating McCullagh's argument with this more general transformation we find that the set of ancillaries always can be pictured as being on the complex unit circle. Furthermore, under the bilinear transformation the ancillaries transform (1) as a rotation through an angle  $2 \arg(c\theta + d)$  if  $ad - bc > 0$ , and (2) as a reflection through  $\arg(c\theta + d)$  if  $ad - bc < 0$ . Thus we find back in a completely different setting the group discussed in section 5. Moreover, this group is connected to a class of transformations that behave roughly the same way as linear transformations, so one may anticipate that a matrix representation of the group may be of value.

The discussion in this section has been very sketchy. There are many open problems if one wants to take this line of attack seriously, which I indeed think we should. A very general open problem is to find a connection to the structural inference of Fraser (1968), which is based upon grouptheoretical arguments and has strong lines back to the fiducial arguments of Fisher (see Hora and Buehler, 1966).

## 13 Discussion

In the last decade we have seen a very interesting exchange of ideas between modern physics and statistics. In a broad review article, Malley and Hornstein (1993) discuss how standard statistical inference has to be amended to be able to handle models formulated in terms of quantum probability. The present article has a much more narrow focus, but attempts to ask a somewhat deeper question: Can we see the contour of a common theory from which both the principles of statistical inference and those of quantum theory can be founded? Though there is a long way to go before we can attempt anything like a definite answer on this question, I think that this paper may be used to point out some directions where it may be fruitful to continue the search.

We have claimed throughout this paper that there seems to be no conflicts between ordinary probability calculus and quantum probability in situations where both are applicable, at least as far as the examples studied here are concerned. If we should attempt to generalize, we could perhaps say that quantum probability is typically not applicable for direct calculations in connection to parametric classes of probability models, though there are definitely exceptions (see Malley and Hornstein, 1993). Ordinary probability calculations are ordinarily not applicable when symmetry considerations can be made *over different experimental choices*.

The situations where we can hope to find that models of the quantum mechanical - or irreducible group representation - kind are of relevance, are situations that are symmetrical under a non-abelian group, and in addition one or more of the following holds:

1. The initial state can be specified by a linear relation between model parameters.
2. The set of conclusions of interest can be specified by a set of linear combinations of parameters.
3. There is a constraint on the model which implies that the calculation can be limited to one irreducible representation of the group.

We know that concepts from other parts of physics (Gibbs sampling, simulated annealing etc.) have turned out to be important in modern statistical simulation theory. In a similar way we can hope that concepts and techniques from quantum theory and group representation theory will turn out to be useful for statistical inference theory and methodology.

However, one can also hope for influence in the other direction. From the three section's discussion of the simple permutation group example above, it seems as if the general irreducible representation theory can tend to give very heavy arguments even

in simple cases. Techniques that have been developed for quantum mechanics may seem to be simpler, at least in this setting, and if applicable, ordinary probability arguments seem to be the simplest to use. One may raise the argument that the potential for using ordinary probability calculations has not been explored enough in, say, modern particle physics. However, before addressing this problem seriously, a further clarification of the roles of the different probability concepts seems to be necessary.

To end with a sweeping question, generalizing far beyond the simple examples discussed here: Can we hope for a future where a typical scientist always has the possibility and willingness in principle to use the technique that is most suitable for his problem, not only the method that he is most familiar with? One can indeed debate if this will ever be possible, or even if it will be desirable: One may easily envisage cases where toolboxes are too heavy to carry. Some progress in this direction may nevertheless be of positive value, however.

## Acknowledgements

I am grateful to Paul Hoyer for telling me about the paper by Aage Bohr and Ole Ulfbeck and sending me a reprint of it, and to Hallstein Høgåsen for his comments. Also, I want to thank Richard Gill and Hans Maassen for sending me preprints of their work.

## References

- Aspect, A., J. Dalibard and G. Roger (1982), Experimental test of Bell inequalities using time-varying analysers. *Phys. Rev. Lett.* **49**, 1804ff.
- Bailey, R.A. (1991), Strata for randomized experiments, *J. R. Statist. Soc. B* **53**, 27-78
- Barndorff-Nielsen, O.E. (1995), Diversity of evidence, and Birnbaum's theorem, *Scand. J. Statist.* **22**, 513-522.
- Bell, J.S. (1966), On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**, 447-452.
- Bohr, A. and O. Ulfbeck (1995), Primary manifestation of symmetry. Origin of quantal indeterminacy. *Reviews of Modern Physics* **67**, 1-35
- Diaconis, P. (1988) *Group Representations in Probability and Statistics* IMS Lecture Notes. Monograph Series, Vol. 11, Hayward, California
- Dirac, P.A.M. (1930) *The Principles of Quantum Mechanics*, Clarendon, Oxford
- Einstein, A., B. Podolsky and N. Rosen (1935), Can quantum-mechanical description of reality be considered complete? *Phys. Rev.* **47**, 777-780

- Everett, H. III (1973) 'Relative state' formulation of quantum mechanics. In: *The Many-Worlds Interpretation of Quantum Mechanics* DeWitt, B.S. and N. Graham, ed., Princeton University Press, Princeton, New Jersey.
- Fraser, D.A.S. (1968) *The Structure of Inference*, Wiley, New York
- Gill, R. and Barndorff-Nielsen, O.E. (1996) Book on quantum mechanics to appear
- Gudder, S.P. (1988) *Quantum Probability*, Academic Press, Boston
- Hamermesh, M. (1962) *Group Theory and its Application to Physical Problems*, Addison-Wesley, Reading, Massachusetts
- Helland, I.S. (1990) Partial least squares regression and statistical models. *Scand. J. Statist.* **17**, 581-607
- Helland, I.S. (1995a) Discussion of: Barndorff-Nielsen, O.E., Diversity of evidence, and Birnbaum's theorem, *Scand. J. Statist.* **22**, 515-519
- Helland, I.S. (1995b) Simple counterexamples against the conditionality principle. *The American Statistician* **49**, 351-356
- Helland, I.S. (1996a) A population approach to the analysis of variance. Submitted *Scand. J. Statist.*
- Helland, I.S. (1996b) Relevance, prediction and interpretation for linear models with many predictors. *Report no.5/1996. Statistical Memoir Series, University of Oslo*
- Helland, I.S. and T. Almøy (1994) Comparison of prediction methods when only a few components are relevant. *J. Amer. Stat. Assoc.* **89**, 583-591
- Hora, R.B. and R.J. Buehler (1966) Fiducial theory and invariant estimation. *Ann. Math. Statist.* **37**, 643-656
- Höskuldsson, A. (1996) *Prediction Methods in Science and Technology*, Thor Publishing
- James, G. and M. Liebeck (1993) *Representation and Characters of Groups*, Cambridge University Press, Cambridge
- Kümmerer, B. and H. Maassen (1996) Elements of quantum probability. In: *Quantum Probability Communications*, no. X, ed. Hudson and Lindsay, World Scientific
- Malley, J.D. and J. Hornstein (1993) Quantum statistical inference. *Statistical Science* **8**, 433-457
- Messiah, A. (1969) *Quantum Mechanics, Vol I and II* North-Holland Publishing Company, Amsterdam
- Meyer, P.-A. (1993) *Quantum Probability for Probabilists*, Springer-Verlag, Berlin
- Morgan, J.P., N.R. Chaganty, R.C. Dahiya and M.J. Doviak (1991) Let's make a deal: The player's dilemma. *The American Statistician* **45**, 284-289
- Müller (1989) *Quantum Mechanics; Symmetries*, Springer-Verlag, Berlin

- Nelder, J.A. (1965) The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Royal Soc. London A* **283**, 147-162
- von Neumann, J. (1955) *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton
- von Rosen, D. (1994) PLS, linear models and invariant spaces. *Scand. J. Statistics* **21**, 179-186
- Serre, J.-P (1977) *Linear Representations of Finite Groups*, Springer-Verlag, Berlin
- Tjur, T. (1984) Analysis of variance models in orthogonal designs. *Internat. Statist. Rev.* **52**, 33-81
- Tjur, T. (1991) Analysis of variance and design of experiments. *Scand. J. Statist.* **18**, 273-323
- Wigner, E.P. (1931) *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*, Braunschweig, Vieweg